Chapter 8: Analysis of Data from the Markson Deep Mine Site, Schuylkill County, PA

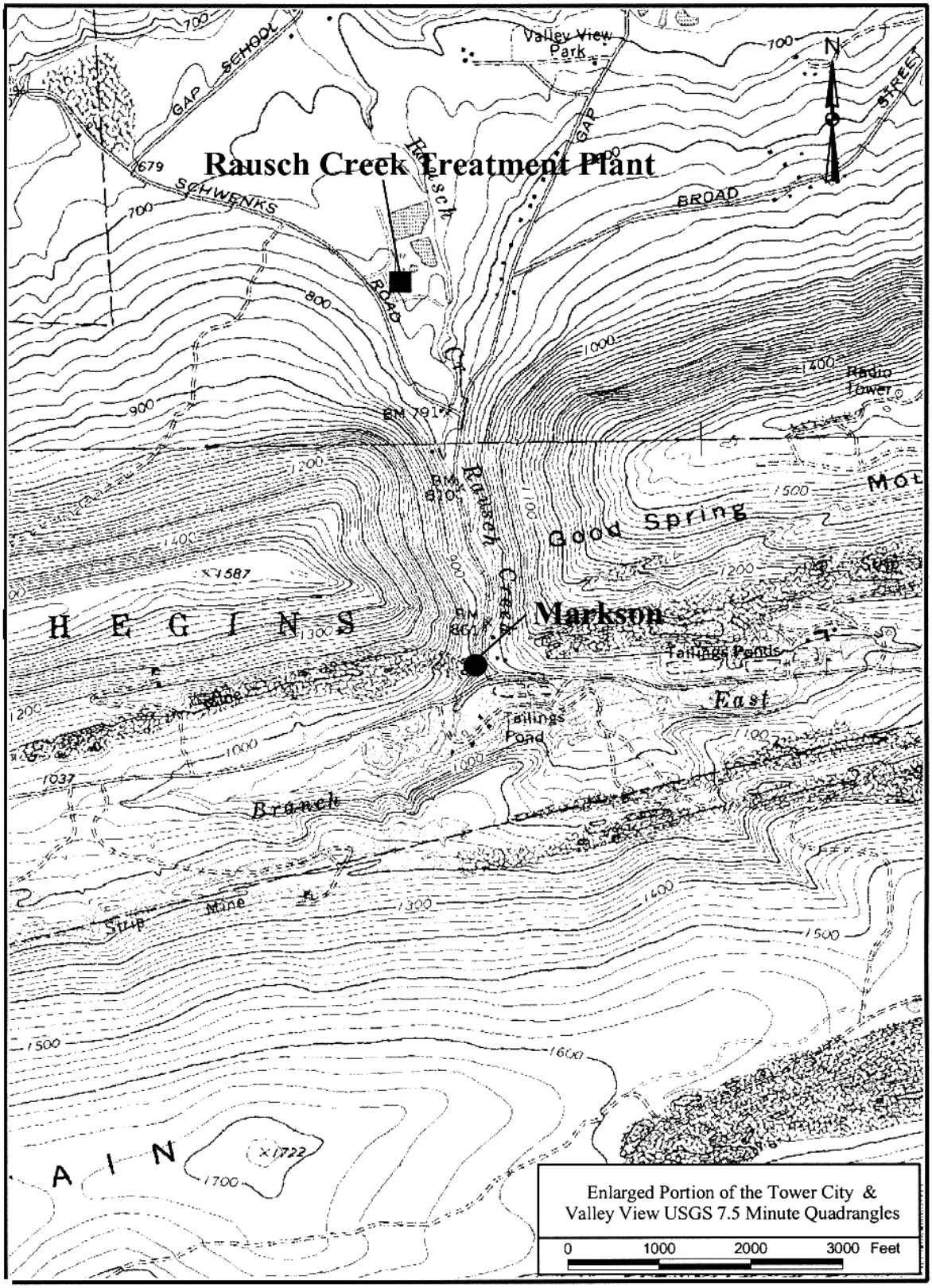
The abandoned Markson Colliery workings are located within the Donaldson Syncline of the Southern Anthracite Coal Field. The Markson discharge is located approximately 1.2 miles upstream from the Rausch Creek Treatment Plant operated by PA DEP in Schuylkill County (Figure 8.0). This discharge emanates from an airway of the abandoned Markson Colliery, and is a principal contributor to the total acid load treated at the plant. The flow and water quality characteristics of the Markson discharge were previously described in Smith (1988), Hornberger et al. (1990), and Brady et al. (1998). The data set used in most of those studies and in this chapter was collected by the Pennsylvania Department of Environmental Protection, Bureau of Abandoned Mine Reclamation (BAMR) which operates the Rausch Creek Treatment Plant. BAMR routinely samples and monitors the Markson discharge and another large abandoned deep mine discharge (Valley View Tunnel discharge) for purposes related to treatment plant operations. Additional data and discussion of the flow and water quality characteristics of the Markson discharge from 1992 to 1999 are contained in Section 5 of the EPA *Coal Remining Statistical Support Document*.

The Markson discharge exhibits water quality characteristics that differ greatly from those of principal discharges from adjacent mines (e.g., the Orchard Airway discharge from the Good Spring No. 1 Colliery and the Tracy Airway discharge from the Good Spring No. 3 Colliery). The pH of the Markson discharge ranges from 3.2 to 3.7, while the pH of the Tracy discharge ranges from 5.7 to 6.5. The distinct chemical differences in two discharges from similar abandoned underground mines in the same coal seams and the same geologic structure are attributable to stratification of large anthracite deep minepools. The Tracy discharge is a "top-water" discharge from a relatively shallow ground water flow system (at elevation 1153 feet), while the Markson discharge emanates from "bottom water" at a much lower elevation (865 feet) in the minepool system. Additional information on the chemical characteristics of stratified anthracite minepools is found in Barnes et al. (1964), Ladwig et al. (1984) and Brady et al. (1998).

The raw data for the Markson discharge are listed in Appendix F. There are 253 observations, and the assembled set is comprised of data on nine parameters as follows: days; flow; pH; acidity; iron; manganese; aluminum; sulfate; ferrous iron. Days were calculated as the number of days between the day an observation was collected and the day the first observation was collected.

The first step was to adjust the data set for missing observations. The first 147 dates had no observations for flow. The original data were, therefore, subdivided into two sets: the first consisting of eight variables (253 observations); and the second consisting of nine variables, including flow (107 observations). A tenth variable (interval) was added by taking the first differences among days to determine the regularity of the intervals between observations.

Figure 8.0: Map of Markson Site



Univariate Analysis

The two data sets (N = 253 and N = 107) were explored initially to determine the shape of the frequency distributions. The variables flow, acidity, aluminum, and ferrous iron were considered to be asymmetric and were transformed to base 10 logarithms. The summary statistics for these data sets following log transformation are shown in Table 8.1 (N = 107) and Table 8.2 (N = 253).

Table 8.1: Summary Statistics (N=107)

	N	N*	Mean	Median	Trimmed Mean	Standard Deviation	Standard Error of the Mean
Intervals	106	1	7.0660	7.0000	7.0208	0.7840	0.0761
Log Flow	107	0	3.1250	3.0892	3.1089	0.1950	0.0188
pH	107	0	3.2458	3.2000	3.2443	0.0954	0.0092
Log Acidity	107	0	1.9960	2.0000	1.9986	0.0743	0.0072
Total Iron	107	0	27.770	27.006	27.680	9.255	0.895
Mn	107	0	4.8971	4.9700	4.8920	0.9498	0.0918
Log Al	107	0	0.2961	0.3191	0.2988	0.1625	0.0157
SO_4	107	0	272.81	271.00	272.41	27.13	2.62
Log Ferrous Iron	105	2	1.3275	1.3979	1.3489	0.2583	0.0250

	Minimum	Maximum	First Quartile	Third Quartile	Coefficient of Variation
Intervals	5.0000	14.0000	7.0000	7.0000	---
Log Flow	2.8791	3.8151	2.9930	3.1978	6.2
pH	3.1000	3.5000	3.2000	3.3000	2.9
Log Acidity	1.5051	2.1818	1.9638	2.0414	3.7
Total Iron	6.900	49.871	21.500	33.762	33.3
Mn	1.2000	8.1600	4.4240	5.3100	19.4
Log Al	-0.1244	0.7686	0.1793	0.3979	54.9
SO_4	210.00	348.00	253.00	292.00	9.9
Log Ferrous Iron	0.5185	1.6758	1.2087	1.5197	75.3

N* = Number of missing data points in data set

Table 8.2: Summary Statistics (N = 253)

	N	N*	Mean	Median	Triggered Mean	Standard Deviation	Standard Error of the Mean
Intervals	252	1	7.333	7.000	7.044	1.910	0.120
Log Flow	107	146	3.1250	3.0892	3.1089	0.1950	0.0188
pH	253	0	3.2362	3.2000	3.2209	0.1508	0.0095
Log Acidity	252	1	2.0515	2.0414	2.0483	0.1123	0.0071
Total Iron	253	0	30.703	28.997	30.410	13.026	0.819
Mn	249	4	5.0439	5.1000	5.0470	0.8351	0.0529
Log Al	246	7	0.34588	0.35622	0.35099	0.14314	0.00913
SO_4	253	0	297.87	293.00	296.84	44.85	2.82
Log Ferrous Iron	241	12	1.3656	1.4393	1.3865	0.3056	0.0197

	Minimum	Maximum	First Quartile	Third Quartile	Coefficient of Variation
Intervals	5.000	28.000	7.000	7.000	---
Log Flow	2.8791	3.8151	2.9930	3.1978	6.2
pH	3.0000	4.4000	3.1000	3.3000	4.7
Log Acidity	1.4771	2.5340	2.0000	2.0197	5.5
Total Iron	3.810	63.500	21.045	39.656	42.4
Mn	1.2000	8.1600	4.6300	5.4200	19.8
Log Al	-0.12436	0.76864	0.28358	0.43497	41.4
SO_4	155.00	510.00	265.00	325.50	15.1
Log Ferrous Iron	-0.1805	1.8037	1.2015	1.5798	22.4

N* = Number of missing data points in data set

The next step was examination of the regularity in the intervals between observations, (i.e., in Table 8.2; the mean is 7.33 days and the median is 7.0 days). The majority of the observations were taken at seven-day intervals as expected. The range was from 5 to 28 days, however, and from the histogram (Figure 8.1a) there were seven intervals of eight days, five intervals of six days, and one interval of five days. The regularity of these observations is desirable yet surprising. It is recommended that determination of the first differences of days should become a routine procedure for long series of observations in order to examine the regularity of sampling intervals.

The next step concerns the magnitude of variation as measured by the coefficient of variation in percent. It should be noted (see Tables 8.1 and 8.2) that when the CV% is calculated [CV% = (σ /mean) *100] after logs have been taken, the mean values tend to be rather low. This will automatically inflate the CV%. It is suspected that this is the case with aluminum and ferrous

iron in both data sets. Most of the variation is very low for flow, pH, and acidity (less than 10%). Only iron, ferrous iron, and aluminum have high CV% values. This is largely due to low mean values rather than high standard deviations.

The frequency distributions of selected variables from these data sets are shown in Figures 8.1a through 8.1i. The number of samples represented in the histograms for each parameter ranges from N = 106 (interval) to N = 253 (sulfate) depending on the number of sample results reported for the corresponding parameter. The histogram for log flow, with N = 107 (Figure 8.1b), is skewed right even after taking logs. The histogram of each data set (N = 252 and N = 107) for log acidity appears fairly regular with some negative skewness after taking logs (Figures 8.1c and 8.1d). The data sets of both iron and manganese result in symmetrical histograms (Figures 8.1e and 8.1f, respectively). Aluminum is symmetrical after taking logarithms (Figure 8.1g) whereas, sulfate is essentially symmetrical without transformation (Figure 8.1h). Ferrous iron is strongly negatively skewed after taking logarithms (Figure 8.1i), making further analysis of this variable suspect.

Figure 8.1a: Histogram of Interval (N=106)

```
Histogram of C10   N = 106   N* = 1
Each * represents 2 obs.

Midpoint    Count
       5        1  *
       6        5  ***
       7       92  **********************************************
       8        7  ****
       9        0
      10        0
      11        0
      12        0
      13        0
      14        1  *
```

Figure 8.1b: Histogram of Log Flow (N=107)

```
Histogram of FLOW   N = 107

Midpoint   Count
     2.9      22  **********************
     3.0      18  ******************
     3.1      33  **********************************
     3.2      15  ***************
     3.3       5  *****
     3.4       5  *****
     3.5       4  ****
     3.6       2  **
     3.7       2  **
     3.8       1  *
```

Figure 8.1c: Histogram of Log Acidity (data set N=252)

```
Histogram of ACID   N = 252   N* = 1
Each * represents 5 obs.

Midpoint   Count
     1.5       2  *
     1.6       0
     1.7       1  *
     1.8       1  *
     1.9      18  ****
     2.0     112  ************************
     2.1      87  ******************
     2.2      19  ****
     2.3       8  **
     2.4       2  *
     2.5       2  *
```

Figure 8.1d: Histogram of Log Acidity (data set N=107)

```
Histogram of ACID   N = 107

Midpoint   Count
    1.50       1  *
    1.55       0
    1.60       0
    1.65       0
    1.70       0
    1.75       0
    1.80       0
    1.85       0
    1.90       9  *********
    1.95      30  ******************************
    2.00      31  *******************************
    2.05      26  **************************
    2.10       8  ********
    2.15       1  *
    2.20       1  *
```

Figure 8.1e: Histogram of Iron (N=253)

```
Histogram of IRON   N = 253
Midpoint   Count
       5       5  *****
      10      14  **************
      15      24  ************************
      20      29  *****************************
      25      41  *****************************************
      30      34  **********************************
      35      33  *********************************
      40      23  ***********************
      45      16  ****************
      50      18  ******************
      55      10  **********
      60       5  *****
      65       1  *
```

Figure 8.1f: Histogram of Manganese (N=249)

```
Histogram of MN   N = 249   N* = 4
Each * represents 2 obs.

Midpoint   Count
     1.0       1  *
     1.5       0
     2.0       1  *
     2.5       2  *
     3.0       2  *
     3.5       7  ****
     4.0      16  ********
     4.5      44  **********************
     5.0      85  *******************************************
     5.5      58  *****************************
     6.0      19  **********
     6.5       6  ***
     7.0       3  **
     7.5       4  **
     8.0       1  *
```

Figure 8.1g: Histogram of Log Aluminum (N=246)

```
Histogram of AL   N = 246   N* = 7
Each * represents 2 obs.

Midpoint   Count
     0.1       3  **
     0.0       7  ****
     0.1      14  *******
     0.2      24  ************
     0.3      69  ***********************************
     0.4      79  ****************************************
     0.5      40  ********************
     0.6       8  ****
     0.7       1  *
     0.8       1  *
```

Figure 8.1h: Histogram of Sulfate (N=253)

```
Histogram of SO4   N = 253
Each * represents 2 obs.

 Midpoint    Count
      160        2  *
      200        2  *
      240       48  ************************
      280       91  *********************************************
      320       60  ******************************
      360       42  *********************
      400        6  ***
      440        1  *
      480        0
      520        1  *
```

Figure 8.1i: Histogram of Log Ferrous Iron (N=241)

```
Histogram of FE++   N = 241   N* = 12
Each * represents 2 obs.

Midpoint    Count
    -0.2        1  *
     0.0        0
     0.2        0
     0.4        0
     0.6        8  ****
     0.8       11  ******
     1.0       24  ************
     1.2       35  ******************
     1.4       65  *********************************
     1.6       74  ***************************************
     1.8       23  ************
```

Bivariate Analysis

The interrelationships among pairs of variables are summarized in Tables 8.3 (N = 107) and 8.4 (N = 253). The critical level of the correlation coefficient for N = 107 is 0.195. Any r less than 0.195 is not significantly different from zero. For the data set with N = 253, any pair with r less than 0.15 is not significantly different from zero. These critical values are inserted above each table. It was found that the degree of association is best measured by the correlation coefficient squared = r^2 %. Due to the large size of the data set, a correlation as low as r = 0.2 can be considered statistically significant. However, a correlation of r = 0.2 yields $r^2 = 0.04$ (i.e., only 4% of the variation of one parameter can be explained by the other parameter). Therefore, a higher r = 0.4 (r^2 = 16%) was arbitrarily chosen to determine the lower limit of interest.

Table 8.3: Correlations Among Variables (N = 107)

r < 0.195 not significantly different from 0									
	Days	**Log Flow**	**pH**	**Log Acidity**	**Iron**	**Mn**	**Log Al**	**SO_4**	**Log Ferrous Iron**
Log Flow	0.422								
pH	0.536	0.263							
Log Acidity	-0.396	-0.259	-0.191						
Iron	-0.174	-0.610	-0.075	0.211					
Mn	0.385	0.048	0.169	0.089	0.183				
Log Al	0.195	0.474	-0.027	-0.036	-0.402	0.584			
SO_4	-0.003	-0.165	-0.087	0.264	0.359	0.322	0.095		
Log Ferrous Iron	-0.338	-0.690	-0.175	0.380	0.770	0.036	-0.424	0.400	
Interval	-0.141	-0.001	-0.119	0.016	-0.067	-0.064	0.042	0.010	0.008

Table 8.4: Correlations Among Variables (N = 253)

r < 0.15 not significantly different from 0									
	Days	**Log Flow**	**pH**	**Log Acidit**	**Iron**	**Mn**	**Al**	**SO_4**	**Log Ferrous Iron**
Log Flow	0.422								
pH	-0.002	0.263							
Log Acidity	-0.348	-0.259	0.016						
Iron	-0.401	-0.610	0.180	0.124					
Mn	-0.036	0.048	-0.030	0.174	0.177				
Log Al	-0.227	0.474	-0.069	0.133	-0.199	0.435			
SO_4	-0.589	-0.165	0.171	0.322	0.455	0.289	0.116		
Log Ferrous Iron	-0.334	-0.690	0.079	0.140	0.803	0.042	-0.268	0.356	
Interval	-0.100	-0.002	-0.088	0.103	-0.020	-0.058	0.027	0.030	0.018

While it is advantageous to use r^2 rather than r as an indication of the strength of any association between two variables, it is also advisable to examine scatter diagrams to see the relationship graphically displayed. Because r and r^2 are really measures of linear association, in those cases where r is low (and r^2 therefore very low), the graph may show a close curvilinear association. For example, in Figure 8.2a (sulfate vs. flow), there could be a curvilinear inverse relationship between the variables although the scatter at high flows (> 3.4) tends to mask it. Both total iron and ferrous iron are also negatively associated with flow (r^2 = 37% in both data sets, see Figure 8.2b, for example for total iron).

Manganese and aluminum show positive association in both data sets (r^2 = 34% and 16% for the 107 and 253 sample data sets respectively, Figure 8.2c). Apart from flow and manganese, there does not appear to be any meaningful relationship between aluminum and any other variable (Figure 8.2d).

Manganese and flow do not show a significant positive (i.e., r = 0.048) or inverse relationship (Figure 8.2e). Similarly, acidity appears to have no association with any of the other parameters (e.g., Figures 8.2f and 8.2g).

Figure 8.2a: Plot of Sulfate vs. Log Flow

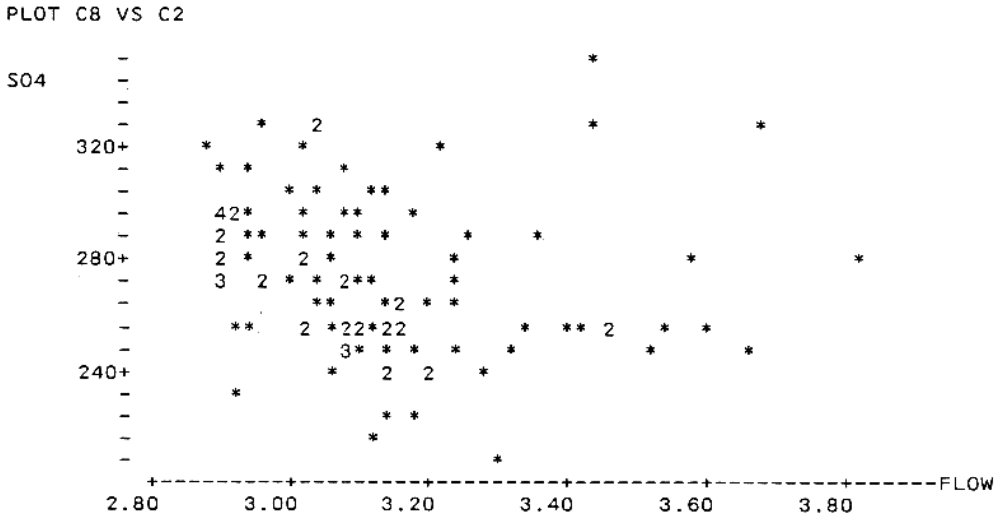


Figure 8.2b: Plot of Iron vs. Log Flow

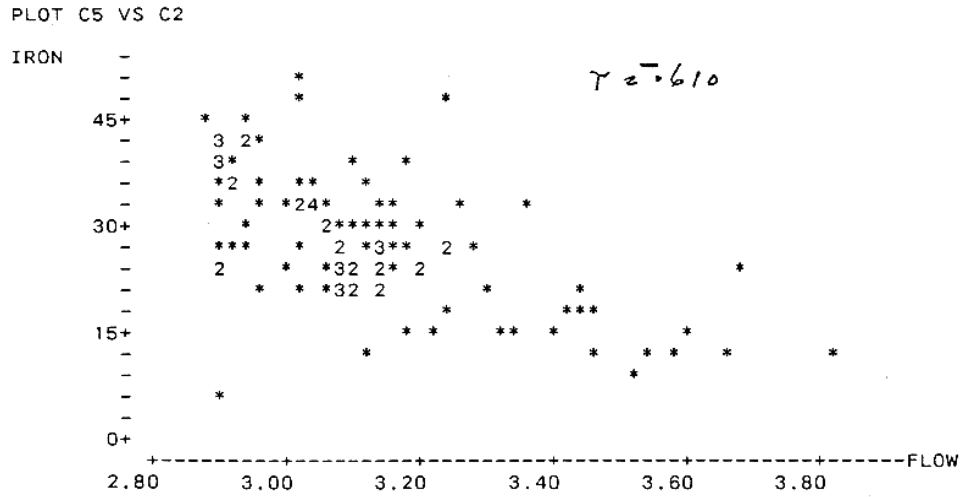


Figure 8.2c: Plot of Manganese vs. Log Aluminum

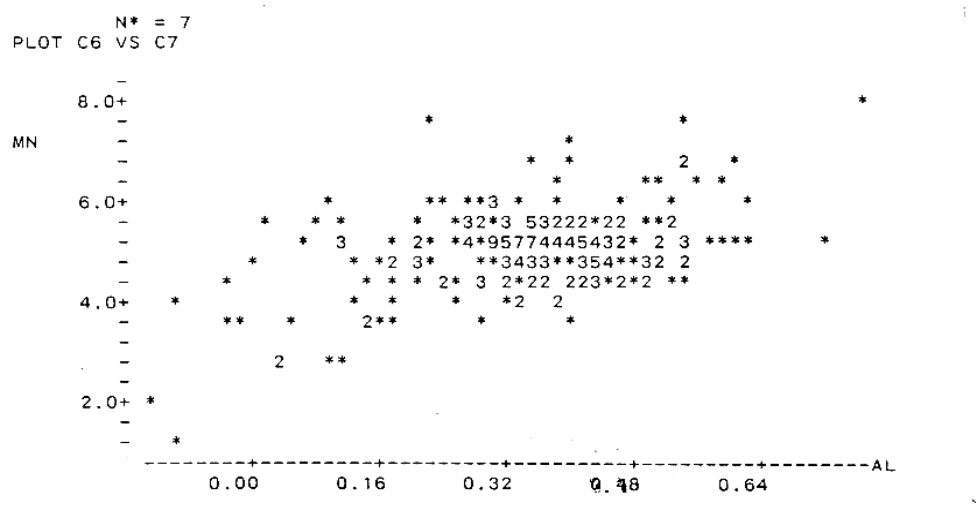


Figure 8.2d: Plot of Log Aluminum vs. Sulfate

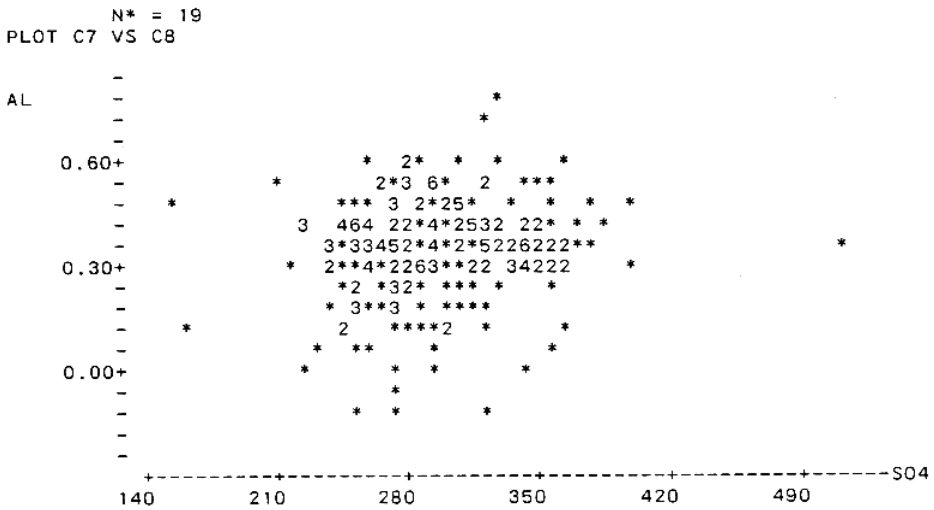


Figure 8.2e: Plot of Manganese vs. Log Flow

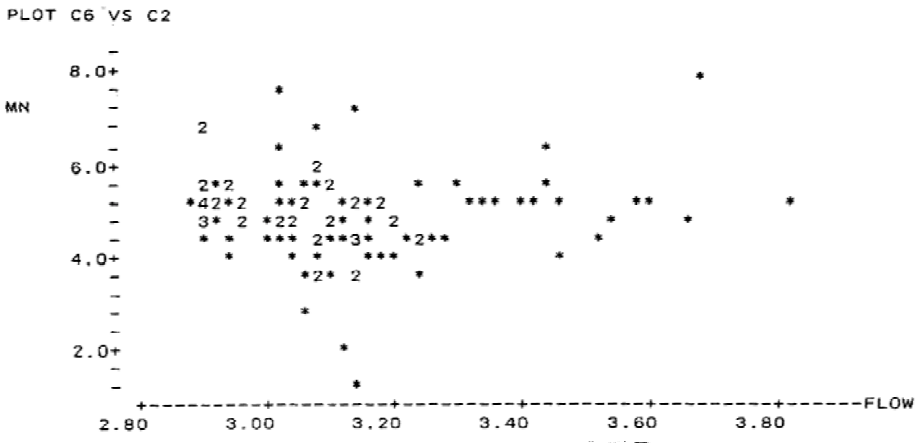


Figure 8.2f: Plot of Log Acidity vs. Log Flow

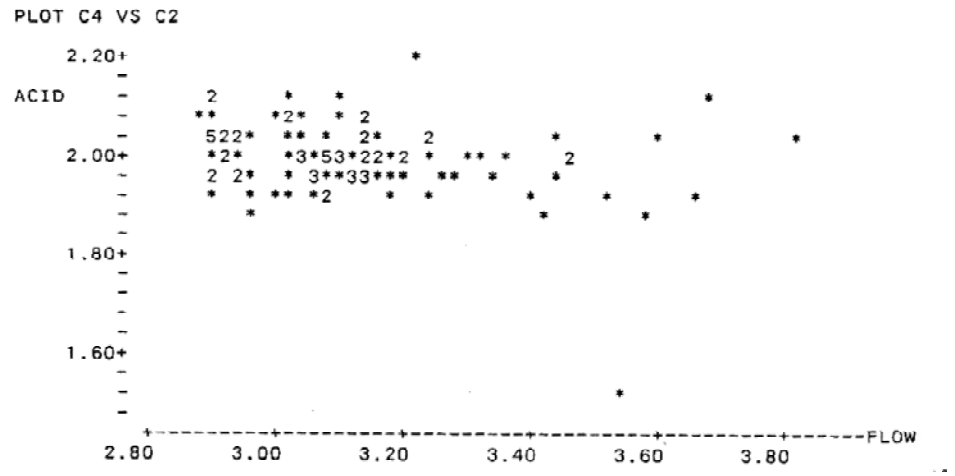
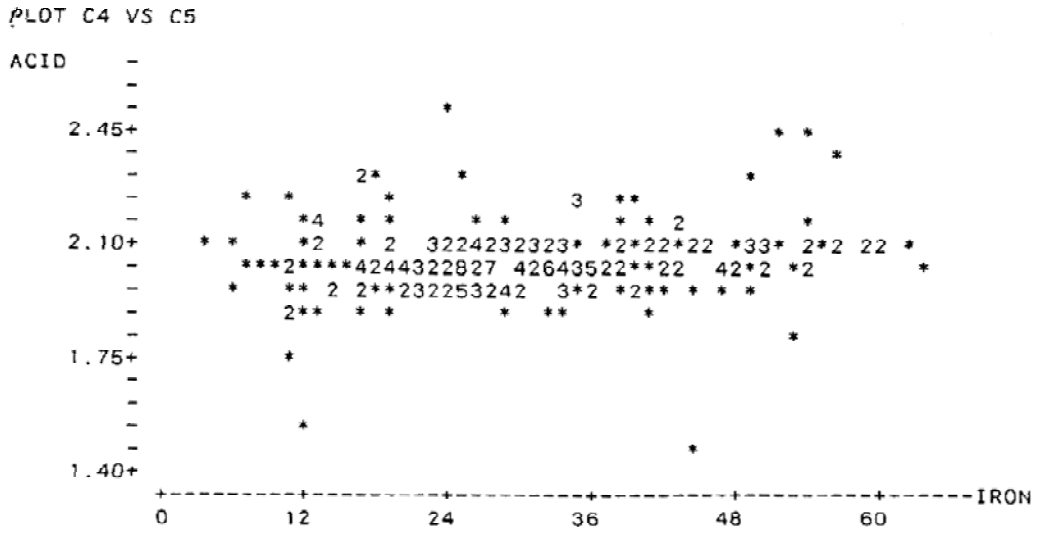


Figure 8.2g: Plot of Log Acidity vs. Iron



Since there does not appear to be any very strong relationship among these variables, they were examined in pairs using cross-correlation functions (CCF). The outcomes are summarized in Table 8.5. There do not appear to be any discrepancies between the correlation coefficient and the cross-correlation coefficient results [i.e., the relationships at zero lag (Tables 8.3 and 8.4) and at any other lags (Table 8.5)]. There are wide regions of the CCF that are above the 0.2 limits demonstrating that, for the most part, interrelationships among these variables are weak to almost non-existent. The only two that stand out are the relationships between flow and iron and iron and ferrous iron (and therefore, between flow and ferrous iron).

Table 8.5: Cross-correlations Among Variables

No.	Variables	r	lag @ r max	$\|r\| >$ 0.2
1.	pH vs. Flow	0.305	-6	- 8 to 7
2.	Acid vs. Flow	- 0.382	-5	- 13 to 0
3.	Iron vs. Flow	- 0.676	-2	- 16 to + 6, 15 to 20
4.	Mn vs. Flow	- 0.334	12	-21, -11, 7 to 20
5.	Al vs. Flow	0.475	0	- 13 to 13
6.	SO_4 vs. Flow	- 0.502	-8	- 18 to - 1, 6 to 20
7.	Ferrous Iron vs. Flow	- 0.740	-1	- 15 to 6, 16 to 20
8.	Acid vs. pH	- 0.321	-5	- 16,- 8,- 7,- 5 to -3, -1, 2, 3
9.	Iron vs. pH	-0.175	-12	none
10.	Mn vs. pH	0.288	13	-14, 6 to 13
11.	Al vs. pH	0.232	-20	- 20
12.	SO_4 vs. pH	0.296	11	8 to 14
13.	Ferrous Iron vs. pH	- 0.227	-13	- 14 to - 12, -10, 2
14.	Iron vs. Acid	0.226	9	0, 4, 9
15.	Mn vs. Acid	- 0.287	13	13
16.	Al vs. Acid	- 0.214	13	13
17.	SO_4 vs. Acid	0.29	2	- 17, - 4 to 0, 2
18.	Ferrous Iron vs. Acid	0.373	0	- 2 to 6, 9, 10
19.	Mn vs. Iron	- 0.283	14	-20, 3, 14, 16 to 18, 20
20.	Al vs. Iron	- 0.562	3	- 20, -9 to 17
21.	SO_4 vs. Iron	- 0.458	20	- 20, -18 to 2, 12 to 20
22.	Ferrous Iron vs. Iron	0.776	0	-11 to 13
23.	Al vs. Mn	0.584	0	-20 to -15, -5 to 1
24.	SO_4 vs. Mn	0.326	-1	-2 to 1
25.	Ferrous Iron vs. Mn	- 0.281	-16	- 20 to - 15, - 10
26.	SO_4 vs. Al	0.419	17	- 20 to - 10, -8, -7, 5, 6, 9 to 20
27.	Ferrous Iron vs. Al	- 0.478	-3	- 16 to 8
28.	Ferrous Iron vs. SO_4	0.473	4	- 20 to - 8, -2 to 13

Time Series Analysis

One of the most striking features of the time series plots (Figures 8.3a through 8.3f) is the limited variation which they show. Another feature of interest is the unusual variation shown by log flow. The major flow event from February 1986 to July 1986 is worth noting and does not appear in any of the other graphs. This emphasizes a lack of relationship between the parameters. Clearly, there appears to be no associated variation among the parameters, except for the pair that includes iron and ferrous iron.

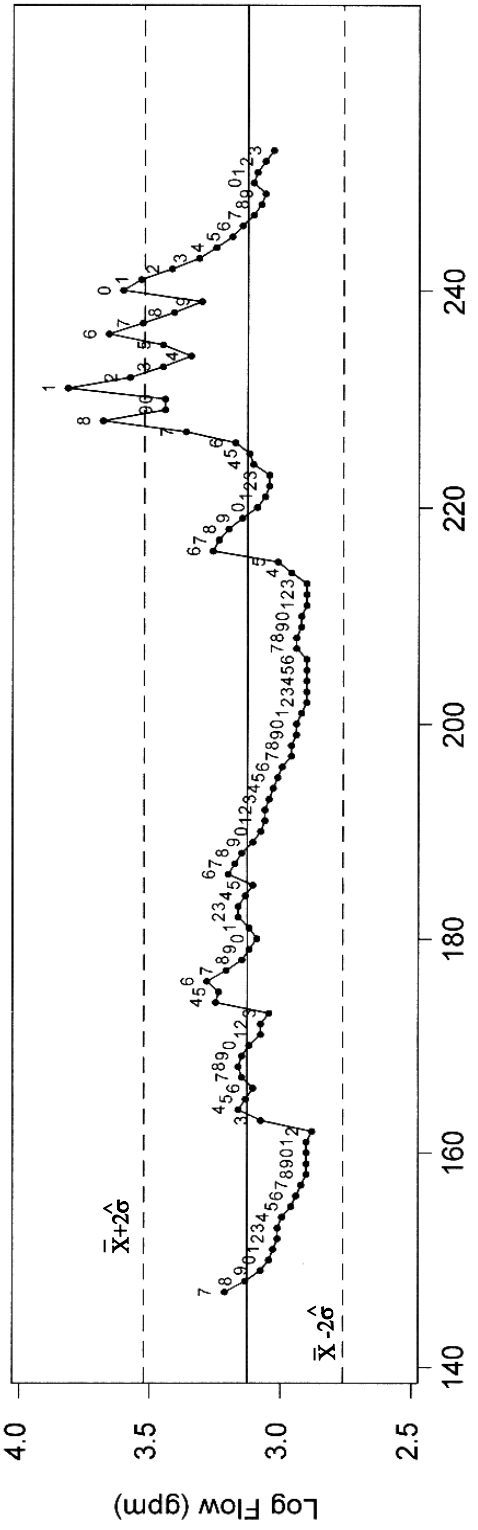


Figure 8.3a: Time series Plot of Flow

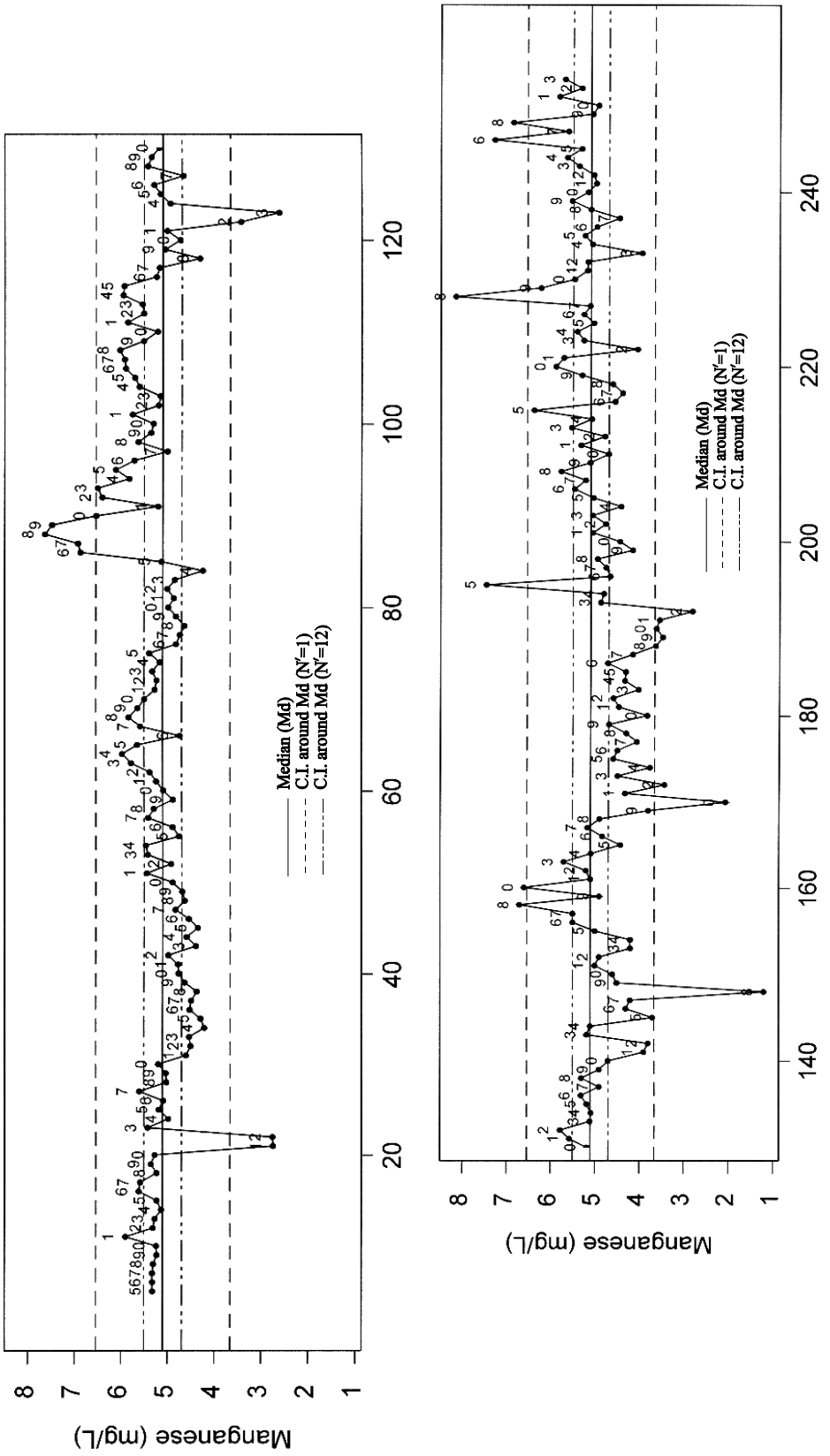


Figure 8.3b: Plot of Manganese

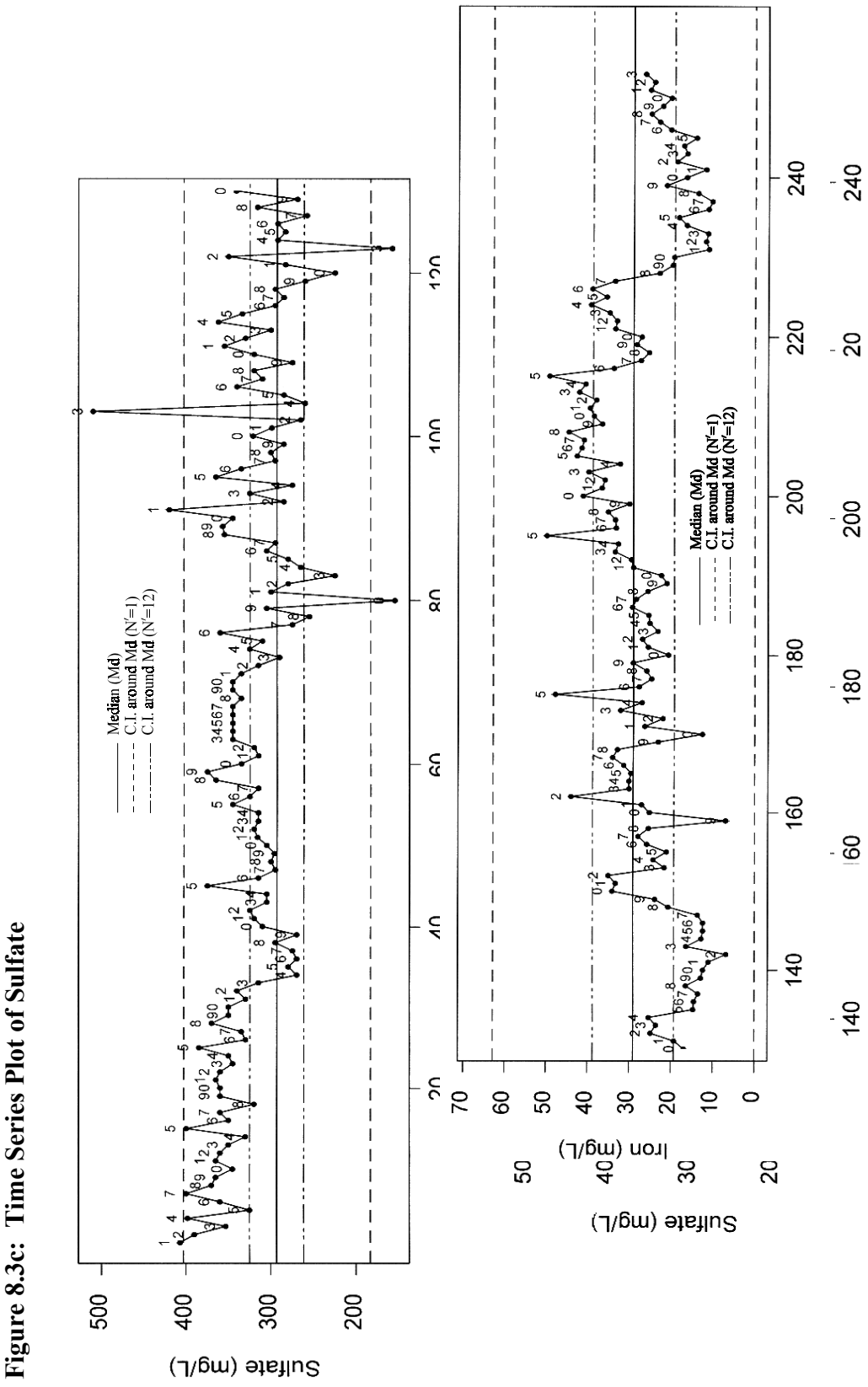


Figure 8.3c: Time Series Plot of Sulfate

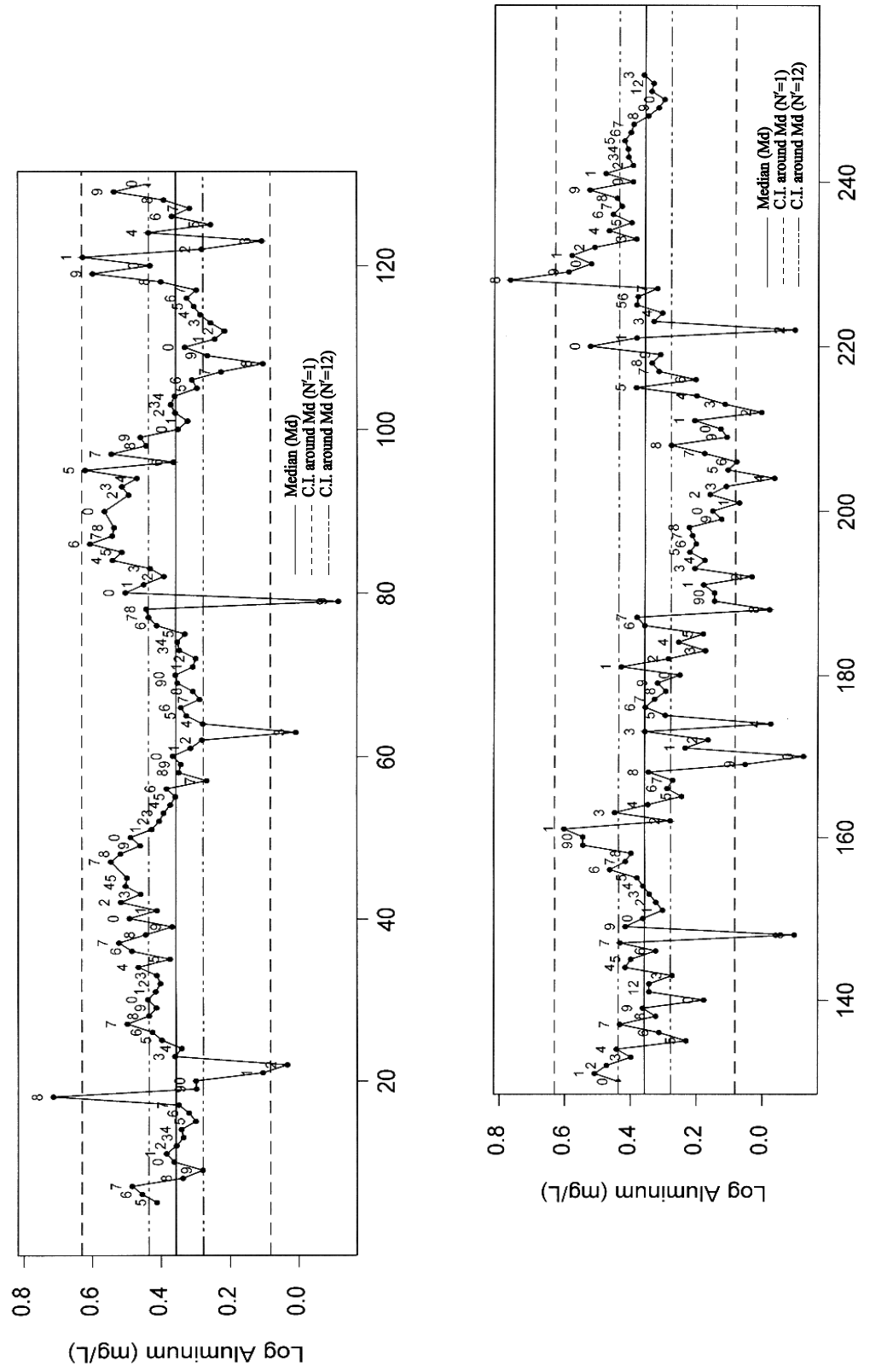


Figure 8.3d: Time Series Plot of Iron

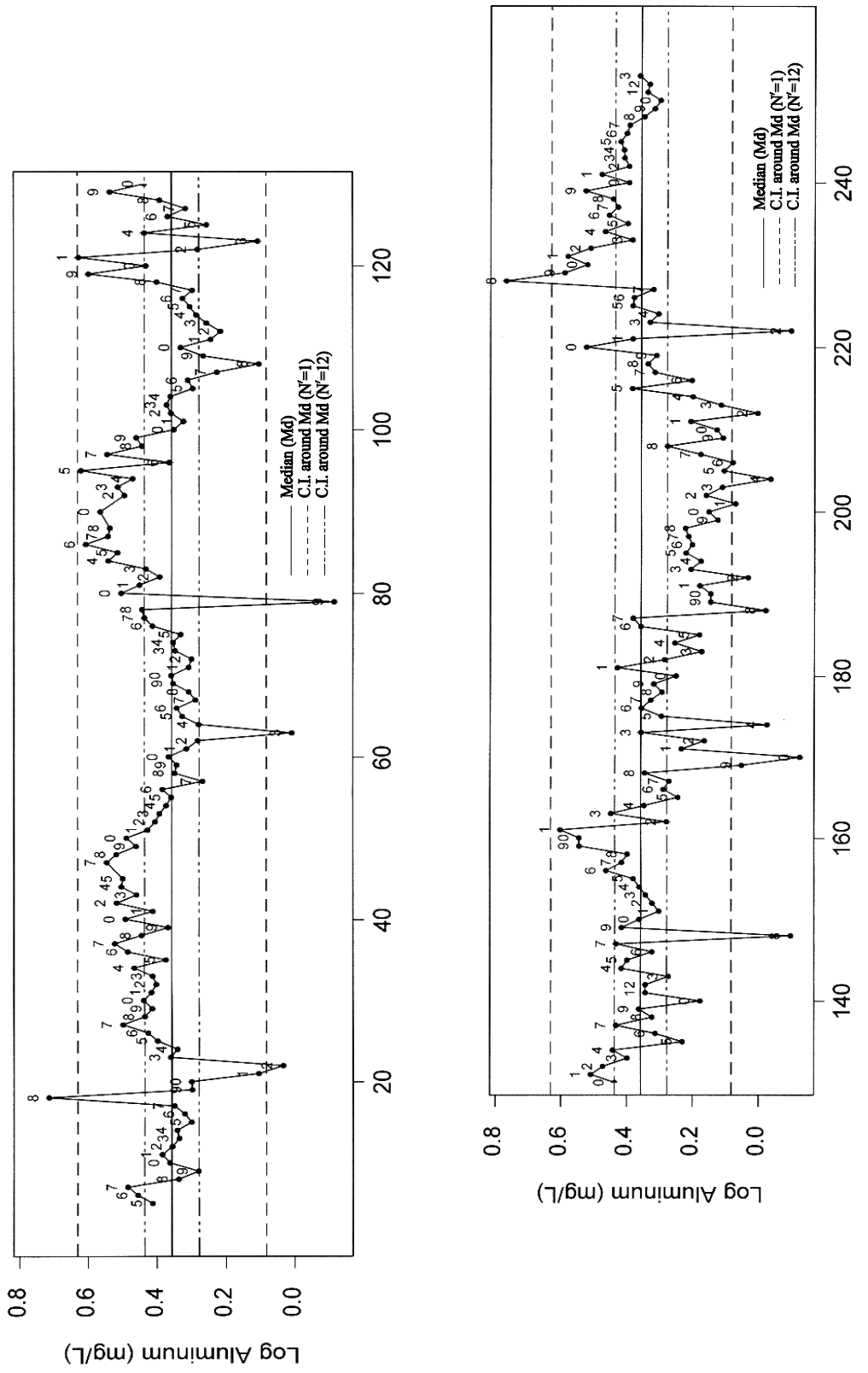


Figure 8.3e: Time Series Plot of Aluminum

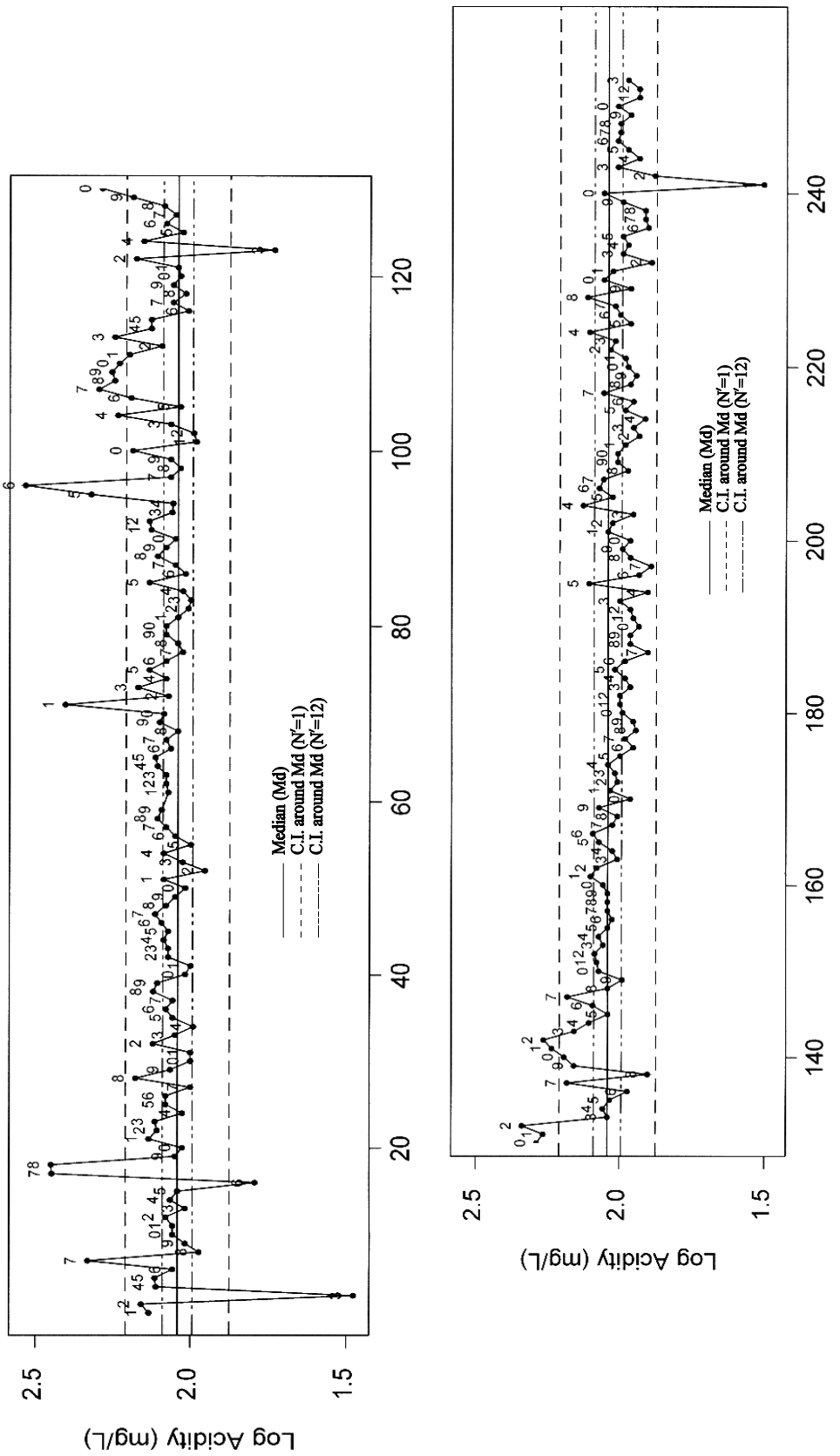


Figure 8.3f: Time Series Plot of Acidity

Quality Control Applied to the Variables

Two measures of quality control are calculated and summarized in Tables 8.6 and 8.7; in Table 8.6 the conventional two standard deviation limits around the mean are given for each of the eight variables (these limits are presented, for example, in Figure 8.3a).

Table 8.6. Quality Control Limits $\overline{X} \pm 2\hat{\sigma}$

Parameter	Mean $\overline{X}$	Standard Deviation $\hat{\sigma}$	$2\hat{\sigma}$	$\overline{X} \pm 2\hat{\sigma}$
Log Flow	3.125	0.195	0.39	2.76 to 3.52
pH	3.246	0.095	0.19	3.056 to3.436
Log Acidity	1.996	0.074	0.148	1.848 to 2.144
Iron	27.770	9.285	18.51	9.20 to 46.34
Mn	4.897	0.950	1.9	2.997 to 6.797
Log Al	0.296	0.163	0.326	(-0.03) to 0.622
SO_4	272.81	27.13	54.26	218.55 to 381.33
Log Ferrous Iron	1.328	0.258	0.516	0.812 to 1.844

Table 8.7. Quality Control Limits: Md ± [1.96 * 1.25(Q_1-Q_3)/1.35 $\sqrt{N'}$] N = 253 data set, except flow, N = 107).

	Median	Q_1	Q_3	Q_3- Q_1	N'	1.35* $\sqrt{N'}$	Lower Limit	Upper Limit
Log Flow	3.089	2.993	3.198	0.205	1	1.350	2.717	3.461
					12	4.677	2.982	3.196
pH	3.2	3.1	3.3	0.2	1	1.350	2.837	3.563
					12	4.677	3.095	3.305
Log Acidity	2.041	2	2.092	0.092	1	1.350	1.874	2.208
					12	4.677	1.993	2.089
Iron	28.997	21.045	39.656	18.611	1	1.350	-4.779	62.773
					12	4.677	19.247	38.747
Mn	5.1	4.63	5.42	0.79	1	1.350	3.666	6.534
					12	4.677	4.686	5.514
Log Al	0.356	0.284	0.435	0.151	1	1.350	0.082	0.630
					12	4.677	0.277	0.435
SO_4	293	265	325.5	60.5	1	1.350	183.204	402.796
					12	4.677	261.305	324.695
Log Ferrous Iron	1.439	1.202	1.58	0.378	1	1.350	0.753	2.125
					12	4.677	1.241	1.637

Furthermore, for Table 8.7, two sample sizes are used, the first with sample size $N' = 1$ and the second with $N' = 12$. The set of quality control limits used in Figures 8.3b through 8.3f are the

limits of a confidence interval (C.I.) Around the median, based on Tukey's non-parametric formula of:

$$\text{Median} \pm [1.96\ (Q_1 - Q_3)\ 1.25 / (1.35\ \sqrt{N'}\ )$$

Two values of N' are used, (namely, $N' = 1$ and $N' = 12$) in Figures 8.3b, 8.3c, 8.3d, 8.3e, and 8.3f.

Many of the observations, which fall outside the limits when $N' = 1$, are single observations and need no activity to explain the exceedance. The longer areas of departure in flow (Figure 8.3a) are due to natural events and presumably, are not related to mining activity. It would be expected that this extreme and long-term departure would be reflected in the variation of the other parameters, but this is not the case.

With nearly all other parameters it appears that the control limits are somewhat tight and that most of the variation outside of the control limits is irregular and of short duration (e.g., manganese in Figure 8.3b and sulfate in Figure 8.3c).

Iron shows two relatively long term, mostly positive deviations beyond the control limits (Figure 8.3d) in the period up to observation 80 (February 20, 1986). These deviations are not repeated in later observations. Sulfate (Figure 8.3c) also extends beyond the upper limits for the first 30 observations (i.e., before February 28, 1985) and appears to decrease with time. The behavior of aluminum within the quality control limits (Figure 8.3e) is similar to manganese and sulfate. However, the aluminum values drop below the lower limit for most of the observations from 185 to 225. Acidity (Figure 8.3f) shows little variation, with a few isolated peaks extending outside the upper limits. In three cases (October 1983, March 1984 and June 1984) consecutive results exceeded the upper limit.

Model Identification

Identification of appropriate models is performed by using the autocorrelation (Acf) and partial autocorrelation (Pacf) functions of the eight parameters. There are three types of functions which can be characterized by appearance. The first type shows a strong steady decline from a high value. Flow (Figure 8.4a), iron (Figure 8.4b), aluminum (Figure 8.4d), and ferrous iron are examples of this type. All these parameters show a large spike at lag 1 in their Pacf (see iron in Figure 8.4c and aluminum in Figure 8.4e). These characteristics imply the parameter possesses a trend which must be removed before further time series analysis. Removal may be achieved by taking first differences. An example can be demonstrated using days, which increase in value throughout the period of observation. Taking first differences results in random walk characteristics (Figure 8.4f). Thus, the first difference is sufficient to make this parameter stationary.

The second type of function shows a less pronounced decline (e.g., pH (Acf in Figure 8.4g)). The Pacf of pH, however, shows a pronounced spike at lag 1, and it too requires taking first

differences to become stationary (Figure 8.4h). Manganese (Acf in Figure 8.4i) is similar to pH except its Pacf does not have a pronounced spike at lag 1 (Figure 8.4j). Therefore, an AR model may be suitable. The first coefficient ($\hat{\Phi}_1$) may suffice for the first difference.

The third type of function is represented by sulfate which appears to show a trend as well as some irregularities (Acf in Figure 8.4k). Before the irregularities can be evaluated the strong spike in the Pacf at lag 1 must be reduced (Figure 8.4l).

Figure 8.4a: Autocorrelation Function of Flow

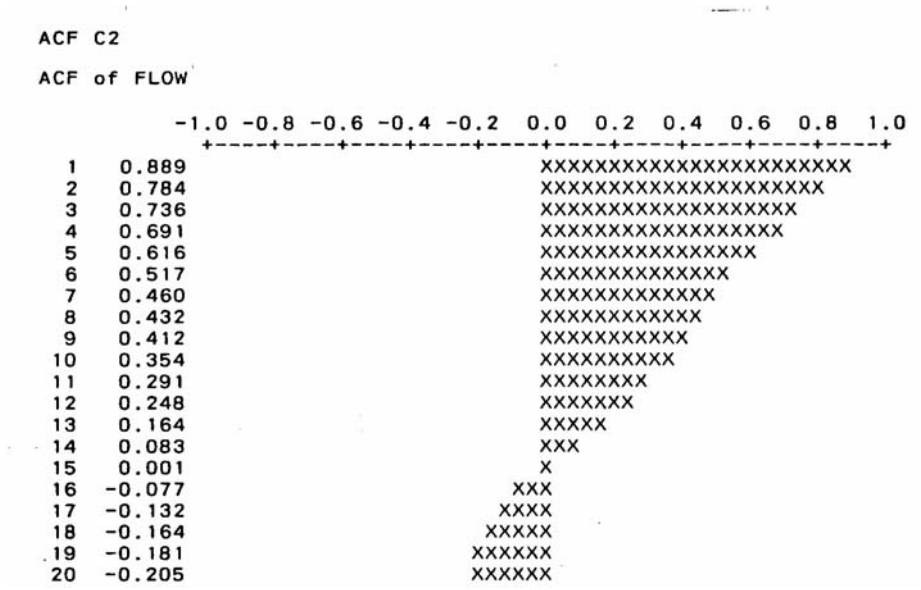


Figure 8.4b: Autocorrelation Function of Iron

```
ACF C5

ACF of IRON

          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1
        +----+----+----+----+----+----+----+----+----+----+----+
  1   0.762                          XXXXXXXXXXXXXXXXXXXX
  2   0.699                          XXXXXXXXXXXXXXXXXX
  3   0.659                          XXXXXXXXXXXXXXXXX
  4   0.603                          XXXXXXXXXXXXXXXX
  5   0.570                          XXXXXXXXXXXXXXX
  6   0.558                          XXXXXXXXXXXXXXX
  7   0.500                          XXXXXXXXXXXXXX
  8   0.469                          XXXXXXXXXXXXX
  9   0.442                          XXXXXXXXXXXX
 10   0.401                          XXXXXXXXXX
 11   0.379                          XXXXXXXXXX
 12   0.327                          XXXXXXXX
 13   0.277                          XXXXXXX
 14   0.200                          XXXXX
 15   0.166                          XXXX
 16   0.101                          XXX
 17   0.045                          XX
 18   0.020                          XX
 19  -0.006                          X
 20  -0.062                        XXX
 21  -0.076                        XXX
 22  -0.115                       XXXX
 23  -0.141                      XXXXX
 24  -0.124                       XXXX
 25  -0.159                      XXXXX
```

Figure 8.4c: Partial Autocorrelation Function of Iron

```
PACF C5

PACF of IRON

          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
        +----+----+----+----+----+----+----+----+----+----+----+
  1   0.699                          XXXXXXXXXXXXXXXXXX
  2   0.245                          XXXXXXX
  3  -0.065                        XXX
  4   0.185                          XXXXXX
  5   0.062                          XXX
  6   0.030                          XX
  7   0.011                          X
  8   0.024                          XX
  9   0.006                          X
 10   0.133                          XXXX
 11   0.090                          XXX
 12  -0.105                       XXXX
 13  -0.015                          X
 14  -0.229                     XXXXXXX
 15  -0.117                       XXXX
 16  -0.034                         XX
 17  -0.055                         XX
 18  -0.051                         XX
 19   0.016                          X
 20   0.021                          XX
```

Figure 8.4d: Autocorrelation Function of Aluminum

```
ACF C7

ACF of AL

          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+
  1   0.516                          XXXXXXXXXXXXX
  2   0.494                          XXXXXXXXXXXXX
  3   0.424                          XXXXXXXXXXXX
  4   0.328                          XXXXXXXXX
  5   0.370                          XXXXXXXXXX
  6   0.235                          XXXXXXX
  7   0.260                          XXXXXXX
  8   0.297                          XXXXXXXX
  9   0.229                          XXXXXXX
 10   0.236                          XXXXXXX
 11   0.212                          XXXXXX
 12   0.174                          XXXXX
 13   0.143                          XXXXX
 14   0.083                          XXX
 15   0.024                          XX
 16   0.022                          XX
 17   0.002                          X
 18   0.019                          X
 19  -0.041                         XX
 20  -0.039                         XX
 21  -0.071                        XXX
 22  -0.051                         XX
 23  -0.078                        XXX
 24  -0.053                         XX
 25  -0.097                        XXX
```

Figure 8.4e: Partial Autocorrelation Function of Aluminum

```
PACF C7

PACF of AL

          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.
            +----+----+----+----+----+----+----+----+----+-
  1   0.511                          XXXXXXXXXXXXXX
  2   0.308                          XXXXXXXXX
  3   0.235                          XXXXXXX
  4   0.126                          XXXX
  5   0.101                          XXXX
  6  -0.152                        XXXXX
  7  -0.094                         XXX
  8   0.096                          XXX
  9   0.040                          XX
 10   0.098                          XXX
 11  -0.028                         XX
 12  -0.032                         XX
 13  -0.077                        XXX
 14   0.025                          XX
 15  -0.086                        XXX
 16   0.026                          XX
 17  -0.017                         X
 18   0.145                          XXXXX
 19  -0.068                        XXX
 20  -0.029                         XX
```

Figure 8.4f: Autocorrelation Function of Intervals

```
ACF C10

ACF of C10

          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+
  1  -0.022                        XX
  2   0.097                         XXX
  3   0.018                         X
  4   0.044                         XX
  5   0.083                         XXX
  6  -0.048                        XX
  7   0.104                         XXXX
  8   0.015                         X
  9  -0.007                         X
 10  -0.026                        XX
 11  -0.065                       XXX
 12   0.066                         XXX
 13  -0.063                       XXX
 14  -0.009                         X
 15  -0.015                         X
 16  -0.034                        XX
 17  -0.012                         X
 18  -0.019                         X
 19  -0.044                        XX
 20  -0.021                        XX
 21  -0.035                        XX
 22   0.001                         X
 23  -0.015                         X
 24   0.011                         X
 25  -0.008                         X
```

Figure 8.4g: Autocorrelation Function of pH

```
ACF C3

ACF of PH

          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+-
  1   0.372                          XXXXXXXXXX
  2   0.290                          XXXXXXXX
  3   0.321                          XXXXXXXXX
  4   0.219                          XXXXXX
  5   0.173                          XXXXX
  6   0.119                          XXXX
  7   0.176                          XXXXX
  8   0.176                          XXXXX
  9   0.108                          XXXX
 10   0.059                          XX
 11   0.078                          XXX
 12  -0.015                         X
 13  -0.054                        XX
 14  -0.074                       XXX
 15   0.019                         X
 16  -0.019                         X
 17  -0.068                       XXX
 18  -0.081                       XXX
 19  -0.036                        XX
 20  -0.016                         X
 21  -0.058                        XX
 22  -0.037                        XX
 23   0.015                         X
 24   0.049                          XX
 25  -0.117                       XXXX
```

Figure 8.4h: Partial Autorcorrelation Function of pH

```
PACF C3

PACF of PH

          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+----+
  1   0.734                          XXXXXXXXXXXXXXXXXX
  2   0.127                          XXXX
  3   0.230                          XXXXXXX
  4  -0.049                           XX
  5  -0.113                        XXXX
  6   0.030                           XX
  7  -0.031                         XX
  8  -0.025                         XX
  9   0.070                          XXX
 10  -0.122                        XXXX
 11   0.003                          X
 12  -0.124                        XXXX
 13  -0.103                        XXXX
 14  -0.012                          X
 15  -0.029                         XX
 16   0.020                          X
 17  -0.172                       XXXXX
 18   0.054                          XX
 19   0.033                          XX
 20   0.253                          XXXXXXX
```

Figure 8.4i: Autocorrelation Function of Manganese

```
ACF C6

ACF of MN

          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+----+
  1   0.523                          XXXXXXXXXXXXXX
  2   0.398                          XXXXXXXXXXX
  3   0.301                          XXXXXXXXX
  4   0.252                          XXXXXXX
  5   0.223                          XXXXXXX
  6   0.208                          XXXXXX
  7   0.198                          XXXXXX
  8   0.140                          XXXX
  9   0.052                          XX
 10   0.009                          X
 11   0.072                          XXX
 12   0.026                          XX
 13   0.092                          XXX
 14   0.049                          XX
 15   0.047                          XX
 16   0.056                          XX
 17   0.109                          XXXX
 18   0.175                          XXXXX
 19   0.171                          XXXXX
 20   0.199                          XXXXXX
 21   0.127                          XXXX
 22   0.187                          XXXXXX
 23   0.128                          XXXX
 24   0.151                          XXXXX
 25   0.157                          XXXXX
```

Figure 8.4j: Partial Autocorrelation Function of Manganese

```
PACF C6

PACF of MN

          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+---+
  1   0.402                          XXXXXXXXXXX
  2   0.232                          XXXXXXX
  3  -0.022                         XX
  4   0.070                          XXX
  5   0.108                          XXXX
  6  -0.021                         XX
  7   0.027                          XX
  8   0.075                          XXX
  9  -0.091                        XXX
 10  -0.100                        XXX
 11   0.126                          XXXX
 12  -0.107                       XXXX
 13   0.079                          XXX
 14   0.035                          XX
 15  -0.004                         X
 16   0.059                          XX
 17   0.127                          XXXX
 18   0.076                          XXX
 19  -0.035                         XX
 20   0.161                          XXXXX
```

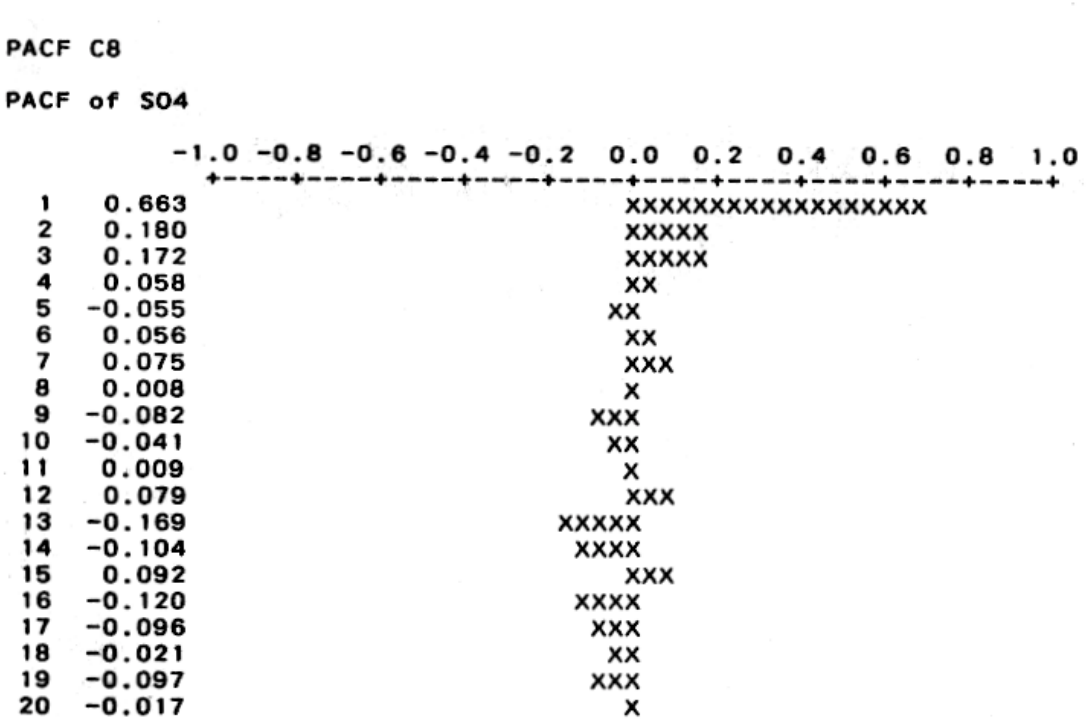
Figure 8.4k: Autocorrelation Function of Sulfate

```
ACF C8

ACF of SO4

          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+----+
  1   0.550                               XXXXXXXXXXXXXX
  2   0.567                               XXXXXXXXXXXXXX
  3   0.584                               XXXXXXXXXXXXXXX
  4   0.505                               XXXXXXXXXXXXX
  5   0.469                               XXXXXXXXXXXX
  6   0.401                               XXXXXXXXXX
  7   0.414                               XXXXXXXXXX
  8   0.392                               XXXXXXXXXX
  9   0.293                               XXXXXXX
 10   0.331                               XXXXXXXX
 11   0.278                               XXXXXXX
 12   0.351                               XXXXXXXXX
 13   0.271                               XXXXXXX
 14   0.296                               XXXXXXX
 15   0.286                               XXXXXXX
 16   0.250                               XXXXXX
 17   0.215                               XXXXX
 18   0.222                               XXXXXX
 19   0.265                               XXXXXXX
 20   0.136                               XXX
 21   0.215                               XXXXX
 22   0.188                               XXXXX
 23   0.128                               XXX
 24   0.174                               XXXX
 25   0.157                               XXXX
```

Figure 8.4l: Partial Autocorrelation Function of Sulfate



Model Fitting to Selected Variables

Since many of the parameters show similar types of variation in terms of their Acf and Pacf representations, only some parameters were submitted to full time series analysis. The outcomes are summarized in Tables 8.8 and 8.9. Table 8.8 summarizes the tests performed on each model fitted to each parameter. Table 8.9 presents the models as equations relating values at time t to previous values of the parameters or its associated shock term (random error term, i.e., a_{t-1}).

There are only 107 observations of flow, and although first differences are indicated by the steep and steady decline of the Acf and the single large spike at lag 1 in the Pacf, it was decided to try an AR (1) model to determine whether the AR coefficient $\hat{\Phi}_1$ was an adequate proxy for the first difference. When an AR (1) was fitted to this variable, both $\hat{\Phi}_1$ and the mean were independent (see r = 0 in No. 1, Table 8.8). The chi-squared value has a probability between 0.20 and 0.10 of not being different from white noise. There were many spikes left in the Acf of the residuals but an AR (1) would suffice as a first approximation. It seems likely that the MA (0,1,1) would be a superior model.

Two models were fitted to the variation in pH. The first was an AR (1,0,0) model. While parameter estimates were independent (r = 0, No. 2. Table 8.8), the chi-square of the Acf of the residuals was significantly different from white noise. There was only one significant spike at lag 3. The second model was an AR (1,1,1) which could more adequately represent variation in

pH. The AR and MA coefficients were not, however, independent (r = 0.64, No. 3 in Table 8.8). The residual Acf yielded a chi-square value that is not significantly different from that of white noise. This model is indeed adequate to represent the series and there were no significant spikes in the Acf or Pacf of the residuals. The residual standard deviation showed a small improvement over the original standard deviation of the series (0.140 and 0.139 versus 0.151, see Table 8.8).

Table 8.8: Tests of the Different Models for Each Parameter

								Standard Deviation	
N	**Parameter**	**Model**	**r[a]**	**Chi-sq**	**d.f.**	**P**	**Spike**	**Resid.**	**Original**
1.	Flow	AR (1,0,0)	0	29.16	22	0.20 > P > 0.10	many	630.43	931.14
2.	pH	AR (1,0,0)	0	34.85	22	0.05 > P > 0.02	1@3	0.140	0.151
3.	pH	AR/MA (1,1,1)	0.64	25.11	22	0.30 > P > 0.20	none	0.139	
4.	Iron	MA (0,1,1)	0	17.17	23	0.90 > P > 0.80	none	8.069	13.026
5.	Iron	AR (1,1,0)	0	27.68	23	0.30 > P > 0.20	1@2	8.349	
6.	Mn	MA (0,0,1)	0	38.27	26	0.02 > P > 0.01	1@2	0.902	0.835
7.	Mn	AR (1,0,0)	0	28.44	22	0.20 > P > 0.10	1@2	0.875	
8.	Mn	MA (0,1,1)	0	30.79	23	0.20 > P > 0.10	1@10	0.721	
9.	SO_4	MA (0,1,1)	0	41.05	23	0.02 > P > 0.01	@3,6,9,12	33.01	44.85
10	SO_4	AR (1,0,0)	0	108.09	22	<0.001	@1,2,3+	36.96	
11	SO_4	AR (1,0,0)	[3][b]	89.70	23	<0.001	many	46.90	
12	SO_4	AR (1,1,0)	88[c]	32.72	21	0.05 > P > 0.02	@6.9	32.83	
13	SO_4	AR (2,1,1)	> 0.7[d]	32.85	21	0.05 > P > 0.02	@6,9,12	32.72	44.85

a. Correlation among the parameter estimates.

b. This [3] = a seasonal @ lag 3.

c. r_{12} = 0.88; r_{13} =0.4; r_{23} = -0.24

d. All coefficients highly positively correlated (i.e., redundant). >0.7

Acidity was considered to be sufficiently similar to pH and for this reason, not require any special testing from the Acf and Pacf of the original series. From the Acf and Pacf and from previous model fitting, an AR (1,1,1) or the MA (0,1,1) would be likely to adequately represent this variable.

Iron showed the same steep decline as flow as well as the same large spike at lag 1 in the Pacf. For this reason, two models were fitted to variation in this parameter. The MA (0,1,1) easily met all tests (see No. 4 in Table 8.8) and is an adequate representative model. The AR (1,1,0) met most of the tests, but not as successfully (see probability for chi-square, No. 5, Table 8.8). There also remained a significant spike at lag 2 in both the Acf and Pacf of the residuals from fitting the AR model. There was a large reduction in the standard deviation compared with its value in regards to the original series (see No. 4, Table 8.8). Ferrous iron was not analyzed because it resembled total iron so closely that similar results would be expected.

Manganese appears to possess a mixture of the characteristics of iron and pH in its Acf and Pacf. For this reason, three models were tried. The simple MA (0,0,1) was not adequate in terms of the chi-square value of the Acf of the residuals (No. 6, Table 8.8). It also possessed a significant spike at lag 2. When a simple AR (1,0,0) model was used, the chi-square of the residuals was not significantly different from that of white noise and the standard deviation of the residuals was reduced from that of the simple MA. There was still a significant spike at lag 2 in both the Acf and Pacf of the residuals.

As a check, a simple MA (0,0,1) was fitted to the first differences of manganese (i.e., an MA (0,1,1)). The results were similar, although the spike at lag 2 disappeared and a weak spike appeared at lag 10. This was determined to be too far along the Acf to be ignored. The standard deviation of the residuals was much improved 0.721 (No. 8, Table 8.8); over 76% improvement over the original series (Table 8.1 wherein the standard deviation is 0.9498 from N = 107).

Sulfate also showed a mixture of types in its Acf and Pacf and for this reason, was explored at greater length; first the MA (0,1,1) was fitted because this model appears to fit many cases in previous reports. The chi-square of the residuals was significantly different from that of white noise. There were many spikes in the Acf and Pacf at lags 3, 6, 9, and 12 implying a seasonal repetition. Next, an AR (1,0,0) was fitted to see how much of the trend shown by the large spike at lag 1 in the Pacf, could be reduced (No. 10, Table 8.8). The chi-square value was very large ($P<<0.001$) and there were many spikes at various lags in the Acf and Pacf of the residuals.

The next model applied was an AR (1,0,0) with a seasonal three term. This model proved ineffective because the chi-square value of the residuals remained very large (probability $<<0.001$ see No. 11, Table 8.8). The next step was to apply AR (1,1,0) with a seasonal AR of lag 2. Lag 2 of the first differences is equivalent to lag 3 in the original series (No. 12 in Table 8.8). This result was a strong improvement in the value of chi-square, but was still significantly different from that of white noise ($0.05>P>0.02$); the Acf and Pacf possessed spikes at lags 6 and 9. Finally, an AR (2,1,1) was applied and all the coefficients ($\hat{\Phi}_1$, $\hat{\Phi}_2$, $\hat{\Phi}_3$) were highly correlated and showed strong redundancies (No. 13, Table 8.8). The chi-square was essentially the same as in the previous model and possessed significantly large spikes at 6, 9 and 12.

It is important at this stage, to examine what the results of these models mean in terms of equations. When a suitable model is found, the coefficients should have some implications of substantive value. In most cases, a simple model possessing an equation that is easy to interpret is adequate. An AR (1,0,0) such as that for flow or pH is an example. As the models become more complicated, interpretation of the equations becomes more difficult. Unless there are definite reasons that a more comprehensive model would be appropriate, it is prudent to use a reasonably simple model.

The equations for the models used in Table 8.8 are summarized in Table 8.9. The first eleven equations are relatively simple. The last two equations, however, are obviously complicated, and it was decided to stop the analysis at this stage. It is suspected that the simple AR or MA models of the first differences are sufficient to represent most of the parameters. However, a seasonal model of some kind is required for sulfate. From the Acf and Pacf of the original

series, an AR model of first differences is likely to be most parsimonious. It will require an additional seasonal term (possibly an AR at lag 3) to remove the remaining significant spikes.

Table 8.9: Model Equations for the Variables (see Table 8.8)

No.	Variable		Equation
1.	Flow	AR(1)	$Z_t = 0.742\ Z_{t-1} + 1487.2 + a_t$
2.	pH	AR (1)	$Zt = 0.372\ Z_{t-1} + 3.235 + a_t$
3.	pH	AR/MA (1,1,1)	$Z_t = 1.121\ Z_{t-1} - 0.121\ Z_{t-2} - 0.808a_{t-1}$
4.	Iron	MA (0,1,1)	$Z_t = Z_{t-1} + a_t - 0.525a_{t-1}$
5.	Iron	AR (1,1,0)	$Z_t = .629\ Z_{t-1} + 0.371_{t-2} + a_t$
6.	Mn	MA (0,0,1)	$Z_t = 4.90 + a_t + .270\ a_{t-1}$
7.	Mn	AR (1,0,0)	$Z_t = 0.404\ Z_{t-1} + 4.913 + a_t$
8.	Mn	MA (0,1,1)	$Z_t = Z_{t-1} + a_t + 0.611\ a_{t-1}$
9.	SO_4	MA (0,1,1)	$Z_t = Z_{t-1} + a_t + 0.725\ a_{t-1}$
10.	SO_4	AR (1,0,0)	$Z_t = 0.550\ Z_{t-1} + 296.9 + a_t$
11.	SO_4	AR (1,0,0)	$Z_t = 0.572\ Z_{t-1} + a_t$
12.	SO_4	AR (1,1,0) (2,0,0)	$Z_t = 1.118\ Z_{t-1} - 0.941\ Z_{t-8} - 1.281\ Z_{t-3} - 0.153z_{t-4} + a_t$
13.	SO_4	AR/MA (2,1,1)	$Z_t = 1.380\ Z_{t-1} + .188\ Z_{t-2} + .182\ Z_{t-3} + a_t - 0.424\ a_{t-1}$

Summary

The first important characteristic of the variables from the Markson site is the lack of wide variation except in the flow variable. The second characteristic is the lack of any strong relationships between pairs of variables. The only high r values are the expected correlations between iron and ferrous iron, flow and ferrous iron, and flow and iron. The iron ferrous iron association is positive, whereas with flow, both are negative (i.e., high flows may lead to dilution of iron). The most striking feature of the time series graphs is the high flow over the period of February 1986 to July 1986, particularly because the flow is in logs. In general, this does not show up in any other variable. There are two very peculiar features which should be emphasized. First, there does not appear to be any reflection of this high flow event in most of the other variables (iron and ferrous iron are exceptions); second, pH shows no relationship to acidity or sulfate.

The most appropriate time series models require a first difference to remove any trend. For many cases, this may be adequately accomplished by an AR (1) term in the models. The residual, after fitting this kind of model, is a close approximation to white noise (i.e., random variation). The residual could, in most cases, be modeled fairly adequately by the usual MA (0,1,1) moving average model. This implies that once the trend is removed, the remaining variation is similar to a random walk. This could account for whatever relationships there are

between the variables, and suggests that linear correlation is not adequate to evaluate the relationships that do exist between the variables. The trend could well be due to the extreme event in the flow variable.

Sulfate however, does appear to show some possible indications of a seasonal pattern. It seems to possess some irregularities which go beyond the "random walk" type of residual. A number of models were tried in this case but none did any better than the MA. Nevertheless, there were many spikes in the Acf of the residual from the first differenced series at what appear to be regular intervals of lags 3, 6, 9, and 12. This implies a seasonal structure at three period intervals (which is a four period interval in the original series). The more complex models failed one or more of the test criteria and rather than complicate the issue further, the analysis was terminated.